

Trabajo Fin de Grado

Explainable Boosting Machine en la predicción de la conversión desde deterioro cognitivo leve a Alzheimer usando datos longitudinales

Explainable Boosting Machine in the prediction of Alzheimer's disease conversion from Mild Cognitive Impairment using longitudinal data

Autor

Pablo Cervera López

Directoras

Mónica Hernández Giménez
Elvira Mayordomo Cámara

Grado en Ingeniería Informática
Escuela de Ingeniería y Arquitectura

2024

Resumen

El Alzheimer es una enfermedad neurodegenerativa crónica que constituye la forma más común de demencia, representando entre el 60 % y el 70 % de los casos. Esta enfermedad afecta principalmente a personas mayores de 65 años, aunque también puede presentarse en personas más jóvenes, conocido como Alzheimer de inicio temprano (Late Onset Alzheimer 's Disease, LOAD). Desafortunadamente, a día de hoy no se conoce una cura y únicamente existen tratamientos para ralentizar el progreso de la enfermedad.

Actualmente, no se conoce un método de diagnóstico preciso para esta enfermedad que no involucre una biopsia post-mortem. Los pacientes diagnosticados con deterioro cognitivo leve (Mild Cognitive Impairment, MCI) pueden terminar desarrollando Alzheimer. Por ello, la detección temprana de un declive cognitivo en este tipo de pacientes es crucial para diagnosticar la enfermedad con antelación y aplicar los tratamientos disponibles, los cuales son más efectivos durante las etapas tempranas de la enfermedad, de manera que se mejore la calidad de vida de los pacientes y sus familias.

Por otro lado, en los últimos años múltiples técnicas de Machine Learning y de aprendizaje profundo (Deep Learning, DL), se han utilizado de forma extensiva para predecir la conversión de un paciente con MCI a Alzheimer. No obstante, estos modelos complejos tienden a sacrificar la interpretabilidad en favor de la precisión, lo que dificulta su aplicación al ámbito clínico. Muchas veces, los expertos en el campo de la medicina se muestran reacios a confiar en el resultado de los métodos denominados black-box debido a que carecen de explicaciones completas y fáciles de entender de sus decisiones.

En este trabajo se ha utilizado un método de Machine Learning explicable, en concreto Explanaible Boosting Machine (EBM), en el problema de predicción de la conversión de MCI a Alzheimer en un periodo de tiempo entre uno y cinco años. Este tipo de técnicas proporcionan interpretabilidad mientras mantienen la precisión, un factor fundamental para su aplicación clínica. El estudio se ha centrado en seleccionar las features más relevantes para la predicción y estudiar su relevancia en los modelos resultantes. Se han considerado cinco conjuntos de features utilizados en estudios previos, todos ellos incluyen datos de distintas modalidades obtenidos de la base de datos ADNI (Alzheimer 's Disease Neuroimaging Initiative).

Los resultados obtenidos muestran que los métodos explicables como EBM no solo obtienen un buen rendimiento, sino que también proporcionan insights valiosos sobre las features más influyentes en la progresión de MCI a Alzheimer. Esto puede facilitar una mejor comprensión de la enfermedad y apoyar a los profesionales médicos en la toma de decisiones clínicas. Como conclusión, este trabajo demuestra el potencial de las técnicas de Machine Learning explicable para el diagnóstico temprano del Alzheimer y sugiere que la combinación de precisión e interpretabilidad es crucial para su adopción en el ámbito clínico.

Glosario

LOAD Late Onset Alzheimer's Disease. Alzheimer de Inicio Temprano.

MCI Mild Cognitive Impairment. Deterioro Cognitivo Leve.

DL Deep Learning. Aprendizaje Profundo.

EBM Explainable Boosting Machine.

ADNI Alzheimer's Disease Neuroimaging Initiative.

AD Alzheimer's Disease. Enfermedad del Alzheimer.

MRI Magnetic Resonance Imaging. Imagen por Resonancia Magnética.

PET Positron Emission Tomography. Tomografía por Emisión de Positrones.

CSF Cerebrospinal Fluid. Fluido cerebroespinal.

APOE Apolipoproteína E.

DTI Diffusion Tensor Imaging. Imagen por Tensor de Difusión.

xAI Explainable Artificial Intelligence. Inteligencia Artificial Explicable.

GAM Generalized Additive Model.

ICV IntraCranial Volume. Volumen IntraCraneal.

DX Diagnose. Diagnóstico.

SMOTE Synthetic Minority Over-sampling Technique.

AUC Area Under the receiver operating Curve.

SHAP SHapley Additive exPlanations.

LIME Local Interpretable Model-agnostic Explanations.

FAQ Functional Activities Questionnaire. Cuestionario de Actividades Funcionales.

CDRSB Clinical Dementia Rating Scale Sum of Boxes.

MMSE Mini-Mental State Examination.

FDG Fluoro-deoxyglucose. Fluorodesoxiglucosa.

ADAS Alzheimer's Disease Assessment Scale.

RAVLT Rey Auditory Verbal Learning Test.

PACC Preclinical Alzheimer Cognitive Composite

Índice

1	Introducción	1
2	Métodos	4
2.1	Conjunto de datos	4
2.2	Preprocesado de datos	7
2.3	Modelo de entrenamiento	8
2.4	Explicabilidad	9
3	Resultados	10
3.1	Resultados utilizando los datos de la visita en baseline	10
3.2	Resultados utilizando los datos de la visita en el mes 12	11
3.3	Resultados utilizando dos instantes temporales consecutivos	11
3.4	Explicabilidad global	12
3.4.1	Explicabilidad global del modelo RSF_set para la visita en baseline	12
3.4.2	Explicabilidad global del modelo ADNIMERGE_set para la visita en el mes 12	14
3.4.3	Explicabilidad global del modelo EBM_set utilizando dos instantes temporales consecutivos	14
3.5	Explicabilidad local	17
3.5.1	Explicabilidad local del modelo RSF_set para la visita en baseline	17
3.5.2	Explicabilidad local del modelo ADNIMERGE_set para la visita en el mes 12	19
3.5.3	Explicabilidad local del modelo EBM_set utilizando dos instantes temporales consecutivos	19
4	Discusión	22
5	Conclusiones	27
Anexo A: Resultados complementarios		
A.1	Resultados utilizando la visita en baseline	
A.2	Resultados utilizando la visita en el mes 12	
A.3	Resultados utilizando dos insantes temporales consecutivos	
Anexo B: Test del Wilcoxon		
Anexo C: Explicabilidad global complementaria		
C.1	Explicabilidad global para la visita en baseline	
C.2	Explicabilidad global para la visita en el mes 12	
C.3	Explicabiliadad global utilizando dos instantes temporales consecutivos	
Anexo D: Explicabilidad local complementaria		

1. Introducción

La enfermedad del Alzheimer (Alzheimer's Disease, AD) es una enfermedad neurodegenerativa de deterioro progresivo de las funciones conductuales y cognitivas, como la memoria, el lenguaje, la atención, el razonamiento y el juicio. Actualmente, el Alzheimer es la sexta principal causa de muerte en los Estados Unidos [1]. Asimismo, los datos más recientes indican que, de aquí a 2050, la prevalencia de la demencia se duplicará en Europa y se triplicará en todo el mundo, y esa estimación es 3 veces mayor cuando se basa en una definición biológica (y no clínica) de la enfermedad del Alzheimer [2]. Debido al actual envejecimiento de la población en todo el mundo, surge una necesidad creciente de desarrollar estrategias de tratamiento que permitan controlar potencialmente la demencia en desarrollo. Por otro lado, el deterioro cognitivo leve (Mild Cognitive Impairment, MCI) se asocia a un riesgo mucho mayor de padecer la enfermedad de Alzheimer y a un ritmo más rápido de deterioro de la función cognitiva [3].

Una de las iniciativas más importantes en el ámbito de la enfermedad del Alzheimer es ADNI (<https://adni.loni.usc.edu/>) surgida en 2004, donde los investigadores se unen para la recopilación, validación y uso de datos relacionados con esta enfermedad, como imágenes por resonancia magnética (Magnetic Resonance Imaging, MRI) y tomografía por emisión de positrones (Positron Emission Tomography, PET), marcadores genéticos, pruebas cognitivas, el fluido cerebroespinal (Cerebrospinal Fluid, CSF), etc. Esta plataforma abarca datos y recursos de pacientes norteamericanos incluyendo pacientes con AD, pacientes con MCI y pacientes con cognición normal. Es preciso señalar que para el estudio realizado en este trabajo se han utilizado datos de esta plataforma.

Los factores de riesgo asociados a esta enfermedad han cambiado en la última década y difieren en función del sexo, la raza y la etnia [4]. Asimismo, otros estudios han destacado entre los posibles factores de riesgo, los factores psicosociales como el nivel educativo o la actividad cognitiva así como la interacción social. Por otro lado, la depresión se identifica como un factor de riesgo significativo para el AD, aunque también puede ser un síntoma temprano de la enfermedad en lugar de una causa directa. Una dieta no saludable, el tabaquismo o el consumo de alcohol también incrementarían el riesgo de desarrollar AD, así como la hipertensión en la edad adulta y la variabilidad excesiva de la presión arterial. Por último, los niveles elevados de colesterol total y otros lípidos en sangre también están relacionados con un mayor riesgo de AD, aunque esta asociación puede estar influenciada por la presencia del alelo APOE ϵ 4 [5].

El desconocimiento de los factores de riesgo favorece a la incomprensión existente de esta enfermedad y la dificultad de su detección. Además, aunque no se puede confirmar definitivamente la enfermedad sin realizar una biopsia post-mortem, en los últimos años se han utilizado neuroimágenes junto con modelos de Inteligencia Artificial para el desarrollo de sistemas de ayuda al diagnóstico del Alzheimer. Entre las modalidades de neuroimagen más utilizadas se encuentran, MRI, MRI funcional, tensor de difusión (Diffusion Tensor Imaging, DTI) y PET [6]. El uso de técnicas avanzadas de aprendizaje automático y la combinación de múltiples modalidades de imagen están mostrando resultados prometedores en la mejora de la precisión diagnóstica y la comprensión de la progresión de la enfermedad. No obstante, la aplicabilidad de los sistemas basados en Inteligencia Artificial depende en gran medida de su explicabilidad, siendo aún más crucial esta propiedad en el ámbito clínico. Esta característica permite al usuario final comprender las decisiones del modelo, fomentando la confianza en que las decisiones de la Inteligencia Artificial son correctas y fundamentadas en datos reales [7].

La explicabilidad por tanto, es clave para fomentar la confianza de los médicos en este tipo de métodos porque permiten conocer en que se basan los resultados y predicciones del mismo. Los médicos se muestran menos reacios a utilizar métodos explicables, es decir, en los que se pueda observar qué features o marcadores han influido en la toma de decisiones del método respecto a utilizar métodos muy complejos y más precisos pero cuya precisión perjudique o sacrifique la explicabilidad de los resultados.

Asimismo, la explicabilidad puede lograrse de dos formas: inherente al método (métodos glass-box) o mediante técnicas xAI externas como SHapley Additive exPlanations (SHAP) o Local Interpretable Model-agnostic Explanations (LIME). Estas técnicas permiten interpretar los métodos denominados black-box que, debido a su estructura y mecanismos de aprendizaje complejos, son no interpretables, es decir, no se sabe por qué el sistema toma decisiones o muestra determinados comportamientos. Por otro lado, los modelos glass-box ofrecen visibilidad sobre cómo un sistema de Inteligencia Artificial toma decisiones y hace predicciones, incorporando restricciones de interpretabilidad dentro de su estructura y mecanismos de aprendizaje [8].

En este trabajo se ha utilizado el método EBM (<https://interpret.ml/docs/ebm.html>), conocido como una técnica glass-box debido a su naturaleza de Inteligencia Artificial explicable (Explainable Artificial Intelligence, xAI). La interpretabilidad de este modelo proviene de su propia fórmula matemática, lo que elimina la necesidad de valores adicionales, haciéndolo intrínsecamente explicable.

Para la predicción de la posible conversión de un paciente con MCI a AD en un periodo de seguimiento de cinco años se ha utilizado EBM, basándose en test cognitivos, medidas MRI de partes del cerebro, biomarcadores y valores socios-demográficos como por ejemplo la edad o el género, evaluando su rendimiento. Además, se ofrece la importancia que las features tienen en la predicción en las diferentes etapas de seguimiento y las explicaciones locales en la predicción de conversión de un individuo concreto. Cabe destacar, que en este estudio se han excluido otros tipos de demencia y se ha centrado el estudio en el Alzheimer. Este trabajo se ha inspirado en el estudio realizado en el paper *Cai J, et.al Explainable Machine Learning with Pairwise Interactions for Predicting Conversion from Mild Cognitive Impairment to Alzheimer's Disease Utilizing Multi-Modalities Data. Brain Sci. 13(11):1535 2023*

El código desarrollado en este trabajo se puede consultar en el siguiente enlace:

<https://github.com/PabloCervera/TFG>

Este TFG se ha desarrollado entre los meses de febrero y julio del 2024. En la Figura 1 se muestra un diagrama de Gantt con las distintas tareas desarrolladas



Figura 1: Diagrama de Gantt del proyecto

2. Métodos

El método que se ha utilizado en este trabajo para la predicción de la conversión de un paciente con MCI a AD es EBM (<https://interpret.ml/docs/ebm.html>). Es un método glass-box diseñado con el propósito de conseguir una precisión comparable a los modelos de aprendizaje automático más avanzados. EBM es un modelo aditivo generalizado (Generalized Additive Model, GAM) de la forma,

$$g(E[y]) = \beta_0 + \sum f_j(x_j)$$

donde g es la función de enlace que adapta el GAM a diferentes escenarios como la regresión o la clasificación. EBM presenta algunas mejoras importantes en comparación con los GAM tradicionales. En primer lugar, EBM es capaz de aprender cada función característica utilizando técnicas modernas de aprendizaje automático como bagging (método para generar múltiples versiones de un predictor y utilizarlas para obtener un predictor agregado [9]) y gradient boosting (metodología de conjunto que combina las predicciones de varios modelos simples, de manera que cada uno mejora el rendimiento general del modelo [10], [11]). En segundo lugar, EBM puede detectar e incluir de manera automática términos de interacción por pares, lo que aumenta aún más la precisión manteniendo la inteligibilidad del modelo, de la forma.

$$g(E[y]) = \beta_0 + \sum f_i(x_i) + \sum f_{i,j}(x_i, x_j)$$

2.1. Conjunto de datos

Los datos que se han empleado en este trabajo han sido recopilados en su totalidad de la base de datos ADNI. Se han utilizado muestras de los cuatro estudios de ADNI (ADNI1, ADNI2, ADNI3 y ADNIGO), donde la mayor parte de las muestras provienen de ADNI1 y ADNI2. La muestra inicial utilizada en este trabajo consta de 1062 pacientes, aunque este número disminuye a lo largo del estudio, tal como se ve en la Tabla 1.

DX	Baseline	Año 1	Año 2	Año 3	Año 4	Año 5
MCI	1062	784(87 %)	515(73 %)	396(67 %)	251(64 %)	142(62 %)
AD	-	116(13 %)	194(27 %)	192(33 %)	142(36 %)	88(38 %)

Tabla 1: Información de los participantes en diferentes duraciones de seguimiento.

Cabe destacar que en este estudio se han decidido utilizar cinco conjuntos de features diferentes en pro de obtener los mejores resultados posibles, ya que no se conoce a priori cuál será el conjunto de features que proporcionará al modelo un mejor rendimiento. Es importante señalar que no siempre la mayor cantidad de features implica un mejor rendimiento, a pesar de su clara influencia, ya que entran en juego distintos aspectos como la cantidad de valores faltantes, el posible incremento de la complejidad del modelo o el sobreajuste del mismo. La relevancia de las features en la predicción realizada por el modelo se determinará gracias a la explicabilidad.

Los cuatro primeros conjuntos de features manejados en este trabajo se han basado en las features utilizadas en distintos artículos que aplican modelos de estudios longitudinales del estado del arte [12], [13], [14], [15]. Estos métodos son, el método ODE-RGRU, una versión de redes neuronales recurrentes que integra ecuaciones diferenciales ordinarias; Minimal RNN, una arquitectura simplificada de redes neuronales recurrentes diseñada para capturar patrones temporales esenciales con menor complejidad computacional; EBM, método utilizado en este trabajo; RSF (Random Survival Forest), un método de Machine Learning basado en árboles para el análisis de datos de supervivencia. El quinto conjunto de features incluye todas las features presentes en el conjunto de datos ADNIMERGE, descargado el 28 de septiembre de 2023. En la Tabla 2, se pueden observar las features que incluye cada uno de los conjuntos. Para facilitar la comprensión y referencia a lo largo del documento, se utilizarán alias específicos para cada conjunto de features, en concreto:

- Conjunto de features utilizado en el método ODE-RGRU se denominará ODE-RGRU_set
- Conjunto de features utilizado en el método MinimalRNN se denominará MinimalRNN_set
- Conjunto de features utilizado en el método EBM se denominará EBM_set
- Conjunto de features utilizado en el método RSF se denominará RSF_set
- Conjunto de features contenido en el fichero ADNIMERGE se denominará ADNIMERGE_set

De manera que a cada conjunto de features, en el caso de los cuatro primeros conjuntos, se le ha denominado con el nombre del método que utiliza esas features, y en el caso del quinto conjunto, se le denomina con el nombre del fichero que contiene dichas features.

En concreto se han utilizado features de diversos tipos, desde test cognitivos, medidas volumétricas de las segmentaciones de diferentes partes del cerebro a partir de MRI (el hipocampo, los ventrículos, la corteza entorrinal, la circunvolución fusiforme, el lóbulo temporal mediano así como el cerebro completo) hasta valores socios-demográficos como por ejemplo la edad. El primer tratamiento de los datos se ha realizado utilizando la herramienta KNIME Analytics Platform (<https://www.knime.com>). En esta primera etapa de tratamiento se han seleccionado únicamente aquellos pacientes que presentaban MCI en la primera visita del estudio (baseline).

Por otro lado, se ha realizado la selección de features de las que consta cada uno de los conjuntos utilizados, cabe destacar que para los cinco conjuntos se ha decidido normalizar las features anatómicas, es decir, las medidas MRI dividiendo por el volumen intracraneal (IntraCranial Volume, ICV), debido a que esta medida es bastante consistente a lo largo de la vida de un paciente. Asimismo, la normalización de estas medidas por el ICV es importante para estimar el alcance de la atrofia causada por la enfermedad [16]. Es preciso señalar que la normalización de las medidas MRI por el ICV es una decisión de diseño propia.

ODE-RGRU_set	MinimalRNN_set	EBM_set	RSF_set	ADNIMERGE_set
MMSE	CDRSB	AGE	AGE	AGE
ADAS11	ADAS11	PTGENDER	PTGENDER	PTGENDER
ADAS13	ADAS13	APOE ϵ 4	PTEDUCAT	APOE ϵ 4
Ventricles	MMSE	CDRSB	APOE ϵ 4	FDG
Hippocampus	RAVLT_immediate	ADAS11	FDG	AV45
WholeBrain	RAVLT_learning	ADAS13	ABETA	ABETA
Entorhinal	RAVLT_forgetting	MMSE	TAU	TAU
Fusiform	RAVLT_perc_	FAQ	PTAU	PTAU
MidTemp	forgetting	Ventricles	CDRSB	CDRSB
	MOCA	Hippocampus	ADAS11	ADAS11
	FAQ	WholeBrain	ADAS13	ADAS13
	Entorhinal	Entorhinal	MMSE	ADASQ4
	Fusiform	Fusiform	RAVLT_immediate	MMSE
	Hippocampus	MidTemp	RAVLT_learning	RAVLT_immediate
	ICV		RAVLT_forgetting	RAVLT_learning
	MidTemp		RAVLT_perc_	RAVLT_forgetting
	Ventricles		forgetting	RAVLT_perc_
	WholeBrain		LDELTOTAL	forgetting
	AV45		DIGITSCOR	LDELTOTAL
	FDG		TRABSCOR	DIGITSCOR
	ABETA		FAQ	TRABSCOR
	TAU		Ventricles	FAQ
	PTAU		Hippocampus	MOCA
			WholeBrain	EcogPtMem
			Entorhinal	EcogPtLang
			Fusiform	EcogPtVisspat
			MidTemp	EcogPtPlan
			ICV	EcogPtOrgan
			mPACCdigit	EcogPtDivatt
			mPACCtrailsB	EcogPtTotal
				EcogSPLang
				EcogSPVisspat
				EcogSPPlan
				EcogSPOrgan
				EcogSPDivatt
				EcogSPTotal
				Ventricles
				Hippocampus
				WholeBrain
				Entorhinal
				Fusiform
				MidTemp
				ICV
				mPACCdigit
				mPACCtrailsB

Tabla 2: Features incluidas en cada uno de los conjuntos utilizados

Asimismo, se ha decidido descartar aquellos pacientes que presentaban en todas las medidas MRI valores faltantes, y se ha realizado la codificación de variables categóricas, en particular de la variable del género, utilizando label-encoding que consiste en asignar un número entero único a cada categoría, en este caso solo hay dos categorías, masculino y femenino. Igualmente, algunos conjuntos de features, concretamente EBM_set y ADNIMERGE_set, incluían como feature el diagnóstico en la primera visita del estudio (DX_bl). No obstante, se decidió eliminar esta feature ya a que podría producir sesgos en el modelo debido a su alta correlación con la variable objetivo.

Finalmente, en algunos conjuntos de features existían ciertas features problemáticas, en concreto ABETA, TAU y PTAU, puesto que algunos pacientes presentaban valores precedidos por el símbolo $< \text{ó} >$, debido al desconocimiento del valor exacto en esas features y ante la problemática de la manipulación de los datos en caso de que se estableciera el valor sin el símbolo, se ha decidido suprimir aquellos pacientes que presentaban estos valores en dichas features. De manera que, como resultado de esta eliminación, algunos conjuntos de features, en concreto MinimalRNN_set, RSF_set y ADNIMERGE_set, presentan un menor número de pacientes, exactamente una muestra inicial de 976 pacientes.

2.2. Preprocesado de datos

Una vez realizado el adecuado tratamiento de las features y previamente a entrenar el modelo es necesario realizar un preprocesado de los datos, es decir, aplicar una serie de técnicas para preparar los datos en crudo antes de que sean introducidos en el modelo. Las técnicas aplicadas en esta etapa son la imputación de datos y el balanceo de clases. Durante la etapa de preprocesado de datos se ha llevado a cabo un KFold con cinco folds con la finalidad de dividir el conjunto en datos de entrenamiento y datos de test. Para evitar el data leakage (problema que ocurre cuando la información del conjunto de prueba se filtra accidentalmente en el conjunto de entrenamiento), en cada uno de los folds se realiza tanto la imputación de datos como el balanceo de clases.

En este estudio se convive con la censura de datos, es decir, los datos en una visita concreta no están disponibles para todos los sujetos durante el periodo de observación. Este fenómeno es común en estudios longitudinales como el que se está realizando. La censura de datos puede introducir sesgos y afectar a la validez de los resultados, para lo cual se lleva a cabo una imputación de los valores faltantes. Es preciso señalar que una fuerte presencia de valores faltantes en una feature determinada puede afectar negativamente al modelo debido a la generación de nuevos valores en base a una muestra muy reducida de valores ya existentes. Por tanto, cuando se dan escenarios de este tipo, se ha decidido no incluir este tipo de features dentro del conjunto de datos utilizado. Para la imputación de datos, se ha utilizado la librería de Python *missingpy* y en concreto el método *MissForest*. Esta fase es indispensable para poder realizar tanto el balanceo de clases como el posterior entrenamiento del modelo, ya que en ambos casos, los métodos utilizados no soportan valores faltantes.

Asimismo, en este estudio existe un desequilibrio en las clases, es decir, que las frecuencias entre los posibles valores son muy dispares, dado que se trata de un problema de clasificación binaria, hay una clase que es mucho más frecuente que la otra. Se ha realizado el balanceo de clases para obtener un mejor rendimiento en determinadas métricas cuyo resultado era demasiado bajo en comparación con otras métricas. El balanceo de clases se ha realizado utilizando la técnica over-sampling, en concreto el método Synthetic Minority Over-sampling Technique (SMOTE) de la librería de Python *imblearn.over_sampling*. Esta técnica también se utilizó en el estudio en el que se basa este trabajo [14] ya que también existía un problema de desequilibrio.

Para solucionar el problema del balanceo de clases se pueden aplicar dos técnicas mayoritarias, over-sampling y under-sampling. Se ha optado por utilizar over-sampling en lugar de under-sampling porque este último método descarta muestras de la clase mayoritaria hasta igualar el número de muestras en ambas clases. A pesar de que el under-sampling es eficaz cuando se trabaja con grandes cantidades de datos y el desequilibrio es pequeño, este no es el caso en el estudio realizado. En este estudio, el desequilibrio es relativamente significativo, especialmente en los primeros años, aunque varía según el año observado. El over-sampling, por su parte, genera duplicados de los datos menos frecuentes en el conjunto de datos, es decir, se crean nuevas muestras hasta equilibrar las frecuencias de ambas clases.

Es preciso señalar que en aquellos métodos de carácter aleatorio, se ha establecido la misma semilla aleatoria para garantizar la reproducibilidad de los resultados. Asimismo para ejecutar el fichero de imputación de datos es necesario utilizar la versión de Python 3.8 debido a la compatibilidad con las librerías empleadas en el proyecto.

2.3. Modelo de entrenamiento

Tras realizar el preprocesado de los datos, se realiza el entrenamiento del modelo. Debido a que se quiere predecir la conversión a AD de un paciente con MCI en un periodo de tiempo determinado, se han realizado cinco modelos distintos, uno para cada año de los que consta el periodo de seguimiento estudiado. En cada modelo, es decir, para cada año, se predice si un paciente con MCI desarrollará Alzheimer o no, siendo la variable de respuesta el diagnóstico del estado cognitivo de cada paciente. Se trata de un clasificador binario donde la clase 0 es MCI (no conversión) y la clase 1 AD (conversión). La predicción se realiza en base a los datos del paciente registrados en la primera visita, es decir, al inicio del estudio, aunque se han realizado pruebas con otras configuraciones como se explicará más detalladamente en el apartado de resultados.

El método EBM utilizado en este estudio es la clase *ExplainableBoostingClassifier*. Esta clase pertenece a la librería Python *interpret.glassbox*. Para cada año, es decir, para cada uno de los modelos, se proporcionan los conjuntos de entrenamiento, tanto la variable X (que representa las features y sus valores para cada uno de los pacientes) como la variable Y (que representa la variable objetivo a predecir). Al igual que en el caso del preprocesado de datos, se ha utilizado la misma semilla en el método *train* de la clase *ExplainableBoostingClassifier* (el único en el que se utiliza semilla aleatoria). En este caso, no hay restricciones referentes a las versiones de Python.

Para analizar el comportamiento del modelo y la validez de los resultados obtenidos, se han utilizado diversas métricas que permiten obtener una visión global del rendimiento del modelo utilizado. Utilizando los conjuntos de datos de test se ha calculado accuracy (proporción de predicciones correctas sobre el total de casos), sensitivity (también conocida como recall, mide la capacidad del modelo para identificar correctamente los casos positivos), specificity (mide la capacidad del modelo para identificar correctamente los casos negativos), f1-score (media armónica entre la precisión y la sensitivity o recall del modelo) y Area Under the receiver operating Curve o AUC (Área bajo la curva ROC, que refleja la capacidad del modelo para distinguir entre clases positivas y negativas).

2.4. Explicabilidad

Asimismo, dado que este método tiene la particularidad de ser glass-box, lo cual supone una ventaja para su aplicación, y que ofrece un marco de aprendizaje totalmente interpretable, no es necesario utilizar técnicas para la mejora de la interpretabilidad como SHAP o LIME. EBM permite el análisis de la explicabilidad, tanto global como local, del modelo. En cuanto a la explicabilidad global, permite conocer aquellas features que influyen en la decisión del modelo y cuáles son los marcadores que favorecen la conversión a AD, en caso de que esta se produzca, o el mantenimiento del paciente con MCI. Además, la explicabilidad local se ajusta a cada paciente en particular, permitiendo un análisis *ad hoc* para cada paciente, una propiedad muy importante en el caso de una enfermedad cuyos factores desencadenantes varían tanto entre los pacientes que la padecen.

En el caso de la explicabilidad global, la gráfica utilizada presenta las features del modelo en el eje vertical y la importancia media absoluta (ponderada) en el eje horizontal. Cada barra representa la contribución de una feature específica a las predicciones del modelo. Mientras que para la explicabilidad local, la gráfica muestra en el eje vertical la lista de las features del modelo junto con sus valores específicos para un caso particular, y en el eje horizontal se muestra la contribución de cada feature a la predicción. Las contribuciones pueden ser positivas o negativas, lo que indica si la feature aumenta o disminuye la probabilidad de la clase predicha. Las barras naranjas indican aquellas features que aumentan la probabilidad de la clase predicha (contribuciones positivas), y las barras azules: indican las features que disminuyen la probabilidad de la clase predicha (contribuciones negativas). Además, el Intercept representa el valor base del modelo antes de considerar el impacto de las features específicas del individuo; es decir, la predicción del modelo cuando todas las features tienen un valor de cero. En ambas gráficas se muestran un máximo de 15 features en caso de que el conjunto presente más. En la Figura 2 se observa un ejemplo de explicabilidad global y local.

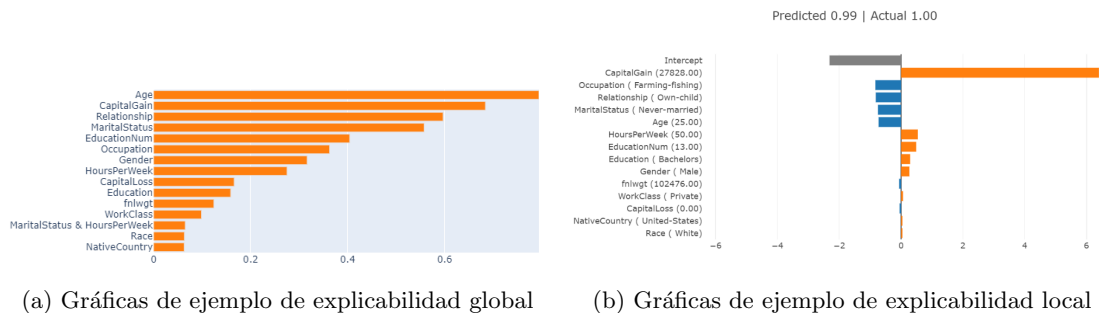


Figura 2: Gráficas de ejemplo de explicabilidad global/local

3. Resultados

Tras haber expuesto los métodos utilizados en este trabajo, en esta sección se presentan los resultados más relevantes de los diferentes experimentos realizados. Cabe destacar que la selección del mejor modelo se basa en la métrica AUC, ya que es una métrica más informativa para la evaluación de la clasificación con datos no balanceados; además, un valor alto indica un buen equilibrio entre sensitivity y specificity, independientemente del desequilibrio en la distribución de las clases. Se ha optado por centrarse en una única métrica para evitar un exceso de información y facilitar la interpretación de los resultados. Los resultados de las métricas restantes para los tres experimentos realizados se incluyen en el Anexo A.

3.1. Resultados utilizando los datos de la visita en baseline

En primer lugar, se analizó el rendimiento del modelo con los diversos conjuntos de features utilizando los datos de los pacientes recogidos en la visita de baseline, es decir, la primera visita del estudio.

El modelo con mejor rendimiento según la métrica AUC es el modelo entrenado con RSF_set, con una media de 87.65 %, como se muestra en la Tabla 3. Destaca especialmente el rendimiento del modelo en los años 2, 3 y 5, donde presenta los mejores resultados. Asimismo se puede observar una importante diferencia entre los resultados del modelo entrenado con ODE-RGRU_set y el resto de modelos, para comprobar si existen diferencias significativas entre los diferentes modelos se ha realizado el test de Wilcoxon, este estudio se encuentra en el Anexo B.

Años	Modelos				
	ODE-RGRU_set	MinimalRNN_set	EBM_set	RSF_set	ADNIMERGE_set
Año 1	75.53 $\pm$ 8.62	83.16 $\pm$ 4.75	82.65 $\pm$ 6.95	83.80 $\pm$ 5.53	82.38 $\pm$ 6.93
Año 2	78.55 $\pm$ 2.82	86.43 $\pm$ 1.88	84.83 $\pm$ 2.66	87.24 $\pm$ 3.20	86.87 $\pm$ 2.16
Año 3	82.23 $\pm$ 4.88	87.63 $\pm$ 3.12	88.18 $\pm$ 3.10	87.52 $\pm$ 3.77	87.42 $\pm$ 3.68
Año 4	85.59 $\pm$ 1.41	89.23 $\pm$ 7.10	91.08 $\pm$ 2.58	90.77 $\pm$ 2.58	90.89 $\pm$ 4.59
Año 5	83.67 $\pm$ 4.28	88.62 $\pm$ 4.09	89.07 $\pm$ 3.58	88.92 $\pm$ 4.55	88.34 $\pm$ 3.64
Media	81.11	87.01	87.16	87.65	87.20

Tabla 3: Media y desviación típica de los distintos modelos para AUC

3.2. Resultados utilizando los datos de la visita en el mes 12

Posteriormente se realizó un experimento en el que se buscaba observar el comportamiento del modelo cuando se utilizan los datos de los pacientes en el mes 12, es decir, pasado un año de la primera visita, para realizar la predicción. De esta manera se pasó a tener cuatro modelos en lugar de cinco, ya que se realiza la predicción en el período de dos a cinco años. Cabe destacar que se eliminó del estudio a aquellos pacientes que habían desarrollado AD en el primer año, es decir, que el diagnóstico en el mes 12 fuera Alzheimer. De manera que en el dataset solo se encontraban los pacientes que tenían MCI tanto en la primera visita como en la visita en el mes 12.

Como se puede advertir en la Tabla 4, el modelo que mejores resultados presenta en AUC es el modelo entrenado con ADNIMERGE_set con un 89.11 % de media, además este modelo es el que mejores resultados obtiene en cada uno de los años por lo que presenta gran robustez en esta métrica en concreto.

Años	Modelos				
	ODE-RGRU_set	MinimalRNN_set	EBM_set	RSF_set	ADNIMERGE_set
Año 2	80.45 $\pm$ 5.11	82.75 $\pm$ 3.48	83.65 $\pm$ 3.30	83.79 $\pm$ 2.77	84.86 $\pm$ 3.44
Año 3	83.51 $\pm$ 6.51	87.09 $\pm$ 3.96	84.73 $\pm$ 2.95	86.97 $\pm$ 3.47	88.25 $\pm$ 3.49
Año 4	86.55 $\pm$ 3.11	88.28 $\pm$ 3.99	89.92 $\pm$ 4.47	88.81 $\pm$ 4.10	90.90 $\pm$ 4.43
Año 5	87.18 $\pm$ 6.43	90.73 $\pm$ 4.02	90.95 $\pm$ 5.94	91.45 $\pm$ 4.32	92.41 $\pm$ 3.80
Media	84.42	87.21	87.31	87.75	89.11

Tabla 4: Media y desviación típica de los distintos modelos para AUC

3.3. Resultados utilizando dos instantes temporales consecutivos

Finalmente, el último experimento realizado fue utilizar conjuntos de features que incluían dos instantes temporales consecutivos, es decir, además de los datos de los pacientes en la primera visita, incluir los datos de los pacientes en la visita del mes 12. De esa manera se aporta más información al modelo.

En este caso, el modelo con mejor resultado promedio para AUC es el entrenado con EBM_set, como se puede observar en la Tabla 5, con un 88.26 % de media. Se destaca especialmente su comportamiento en los dos últimos años, donde su rendimiento es superior al de los otros modelos. Aunque no es el que presenta los mejores resultados en los dos primeros años, su destacado rendimiento en los últimos años le permite ser el mejor modelo en promedio.

Años	Modelos				
	ODE-RGRU_set	MinimalRNN_set	EBM_set	RSF_set	ADNIMERGE_set
Año 2	83.23 $\pm$ 3.75	82.43 $\pm$ 3.98	82.44 $\pm$ 3.37	83.87 $\pm$ 3.09	83.13 $\pm$ 6.19
Año 3	86.17 $\pm$ 5.93	86.91 $\pm$ 4.52	86.78 $\pm$ 1.71	85.25 $\pm$ 6.62	85.99 $\pm$ 4.23
Año 4	90.09 $\pm$ 3.17	88.23 $\pm$ 3.51	91.03 $\pm$ 2.62	89.71 $\pm$ 4.40	88.82 $\pm$ 4.43
Año 5	90.28 $\pm$ 5.74	89.33 $\pm$ 5.43	91.77 $\pm$ 6.22	90.87 $\pm$ 5.19	89.99 $\pm$ 4.70
Media	87.44	86.72	88.26	87.75	86.98

Tabla 5: Media y desviación típica de los distintos modelos para AUC

3.4. Explicabilidad global

Una de las características distintivas de EBM es su capacidad para proporcionar explicabilidad del modelo. La explicabilidad consta de dos partes diferenciadas: la explicabilidad global y la explicabilidad local. La explicabilidad global se ha obtenido utilizando el método *explain_global* y la explicabilidad local con el método *explain_local*, ambos de la clase *ExplainableBoostingClassifier*. Tras analizar los resultados más relevantes de los distintos experimentos realizados, en este apartado se va a proceder a analizar la explicabilidad global del mejor modelo para cada uno de los experimentos, el análisis de la explicabilidad global del resto de modelos se encuentra en el Anexo C. Para evaluar la explicabilidad global de los distintos modelos, se ha analizado el comportamiento de cada uno a lo largo de los diferentes años del periodo de seguimiento (desde el año 1 al 5 o desde el año 2 al 5). Como se ha comentado en el apartado de Métodos, por cada conjunto de features y cada año se ha desarrollado un modelo independiente (de manera que se tienen cinco modelos distintos por cada conjunto de features) y cada uno de estos modelos a su vez, se ha validado utilizando Kfold con cinco folds. Esto implica que se podría analizar la explicabilidad de cada modelo cinco veces por cada año. Por tanto, para simplificar el proceso, se ha decidido utilizar únicamente el quinto fold de cada modelo para el análisis de la explicabilidad.

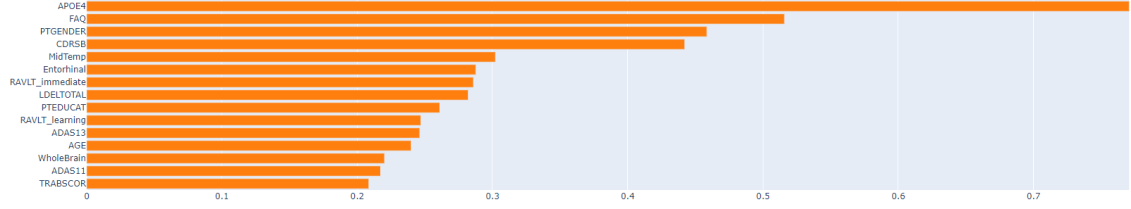
Cabe destacar que se ha comprobado si existe variabilidad en la explicabilidad entre los distintos folds. Cuando se analiza la explicabilidad global para otros folds, existe variabilidad, ya que en cada uno de los folds se seleccionan distintos pacientes, además, como se ha comentado en el apartado 2.2 el preprocesado de datos se realiza en cada uno de los folds de manera independiente. Por tanto, la predicción y por ende la importancia de las features depende de los datos de los pacientes que se utilicen en el entrenamiento del modelo. No obstante, la variabilidad existente se da en las features que presentan una importancia media, y esta no es muy significativa, mientras que las features más importantes no sufren apenas variabilidad entre los folds.

3.4.1. Explicabilidad global del modelo RSF_set para la visita en baseline

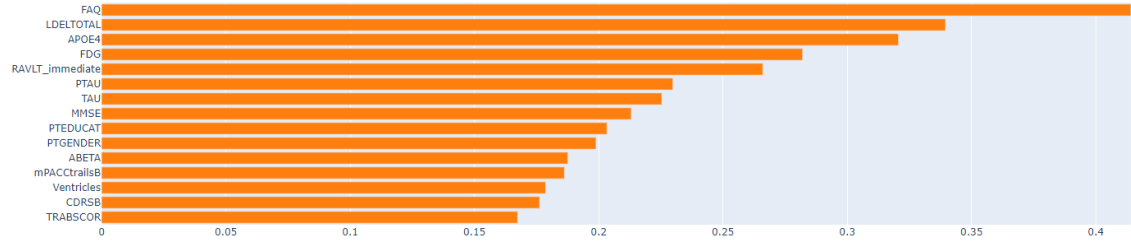
La Figura 3 muestra la relevancia de la feature Functional Activities Questionnaire (FAQ), siendo la más importante en prácticamente todos los años, salvo en el primer año. Otra feature también relevante para la predicción sería el Clinical Dementia Rating Scale Sum of Boxes (CDRSB), presentando una relevancia muy significativa en todos los años salvo en los años 2 y 4. También destaca Fluoro-deoxyglucose (FDG), muy influyente en los 4 últimos años.

Asimismo, tanto APOE ϵ 4 como PTGENDER presentan cierta relevancia. En el caso de APOE ϵ 4, es muy relevante en los cuatro primeros años mientras que en el quinto año apenas es relevante. En el caso de PTGENDER tiene una influencia mediana en todos los años cobrando especial relevancia en los años 1 y 4. Por último, se podría destacar la feature RAVLT_immediative bastante relevante para los tres primeros años aunque en los dos últimos años su relevancia disminuye considerablemente.

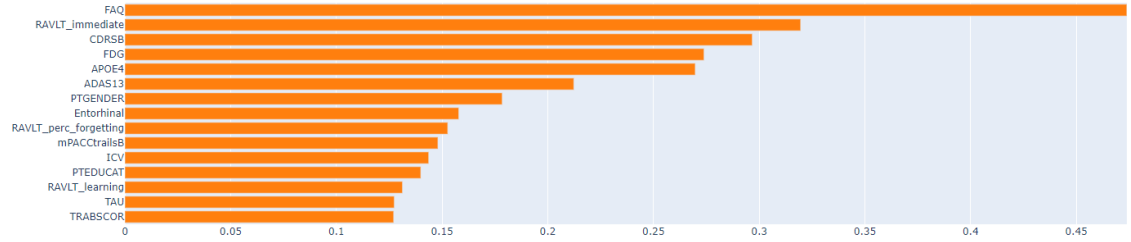
En cuanto a las features anatómicas, se destacan Entorhinal y MidTemp. Entorhinal es relevante en todos los años salvo en el segundo, aunque para los años 4 y 5 su relevancia disminuye respecto al primer y tercer año. Por otro lado, MidTemp es especialmente relevante en los años 1 y 4, mientras que en el resto de los años no es apenas relevante. En resumen, destaca la relevancia de las features FAQ, APOE ϵ 4 y CDRSB, también se destaca la relevancia de FDG, feature asociada a imágenes PET. Asimismo, recalcar la importancia de RAVLT_immediative (test de evaluación neuropsicológica).



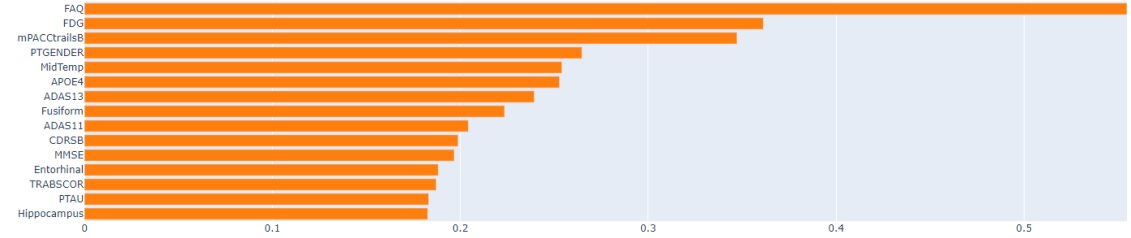
(a) Importancia global en el periodo de seguimiento de un año



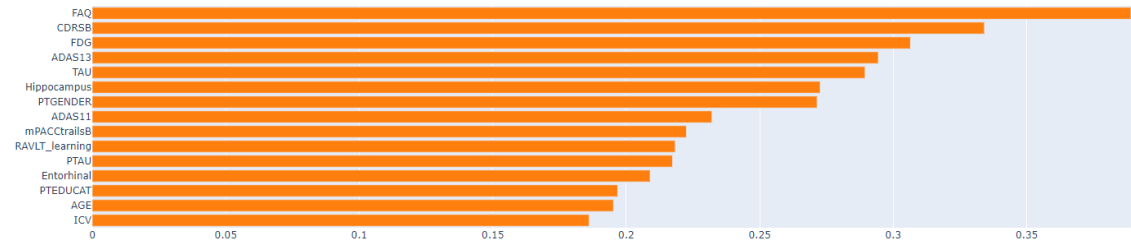
(b) Importancia global en el periodo de seguimiento de dos años



(c) Importancia global en el periodo de seguimiento de tres años



(d) Importancia global en el periodo de seguimiento de cuatro años



(e) Importancia global en el periodo de seguimiento de cinco años

Figura 3: Modelo RSF_set. Explicabilidad global en baseline

3.4.2. Explicabilidad global del modelo ADNIMERGE_set para la visita en el mes 12

Como se observa en la Figura 4, en este caso APOE ϵ 4 es la feature más importante junto con FAQ. APOE ϵ 4 es muy importante en los años 2, 3 y 4, y en el año 5 no es nada relevante. De la misma forma, FAQ también es muy relevante en prácticamente todos los años salvo en el cuarto donde no es nada relevante. Por otro lado, la relevancia de FDG es menor en este caso, ya que únicamente es muy relevante en los años 2 y 5. Algo similar ocurre con CDRSB, el cual es relevante solamente en los años 4 y 5. Por otro lado, RAVLT_immediate es muy relevante en los años 2 y 3 y nada relevante para los dos últimos años. También destaca mPACCdigit cuya importancia es significativa en los años 2 y 3, mientras que en los dos años restantes no es apenas relevante.

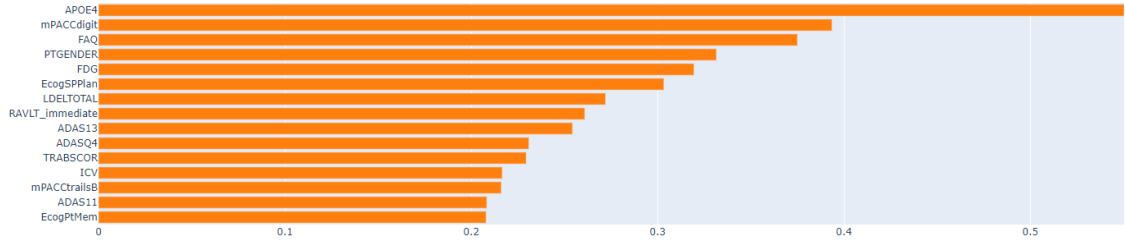
Por otra parte, se destaca la gran variabilidad existente en el cuarto año ya que las features más relevantes varían bastante en relación al resto de años. En este caso, destacan las features EcogPtMem y EgogSPTotal cuya relevancia es muy significativa, pero en el resto de años no tienen apenas relevancia. En cuanto a las features anatómicas, destaca MidTemp, cuya relevancia es muy significativa en los dos últimos años.

Finalmente, se puede concluir que en este conjunto hay mucha variabilidad en la importancia de las features dependiendo del año. Asimismo, las features más relevantes son FAQ y APOE ϵ 4, sin embargo, otras features como FDG y CDRSB son menos relevantes en comparación a lo que ocurría en este conjunto cuando se utilizan los datos de baseline. RAVLT_immediate tiene un comportamiento similar a la visita inicial y destaca la importancia de mPACCdigit con un comportamiento similar a RAVLT_immediate, pero que apenas tenía relevancia en baseline.

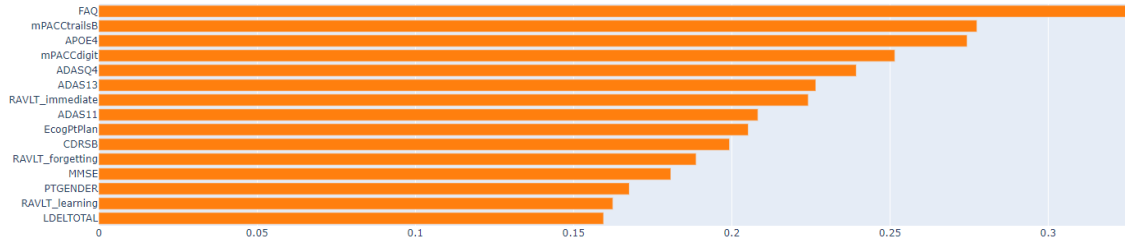
3.4.3. Explicabilidad global del modelo EBM_set utilizando dos instantes temporales consecutivos

Como se observa en la Figura 5, la feature más destacada es FAQ, tanto en el valor de la visita en baseline como en el mes 12. La relevancia de estos valores varía con los años: en el año 2, ambos son muy importantes; en los años 3 y 4, FAQ_m12 es más relevante; y en el último año, FAQ en baseline vuelve a ser más importante. Además, se destaca APOE ϵ 4, cuya importancia general es menor, siendo muy importante en el primer año, pero en los años restantes es poco relevante, aunque muestra un repunte en el cuarto año. También destacan Mini-Mental State Examinatio (MMSE) y ADAS11 (Alzheimer's Disease Assessment Scale), ambos en baseline, con una importancia significativa. El valor de MMSE en el mes 12 solo es relevante en el primer y último año, mientras que ADAS11 es muy poco relevante en el mes 12.

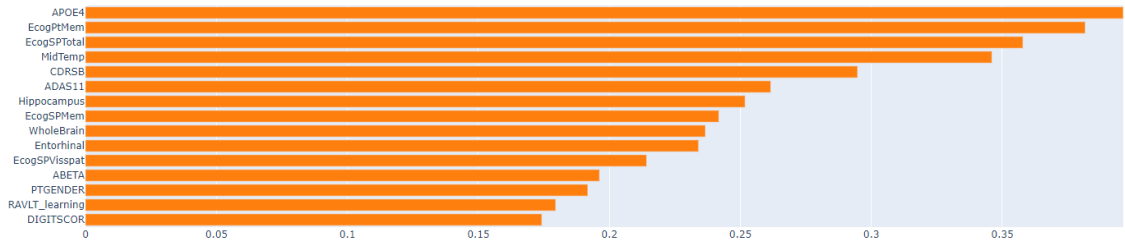
Por otro lado, se destaca CDRSB, cuyo valor en baseline es importante en los tres últimos años, mientras que en el mes 12 solo es relevante en el último año. Un caso similar es el de ADAS13, que es muy relevante, en baseline, en los años 2, 3 y 5, siendo el año 3 el único en el que ambos valores (baseline y m12) son bastante relevantes. Finalmente, en cuanto a las features anatómicas se destaca su escasa relevancia en términos generales, siendo Hippocampus la más relevante, también es preciso señalar la menor importancia de PTGender en la predicción en comparación con la visita inicial. Como conclusión, las features más relevantes son FAQ (tanto en baseline como en m12) y ADAS13, y destaca la menor influencia de APOE ϵ 4.



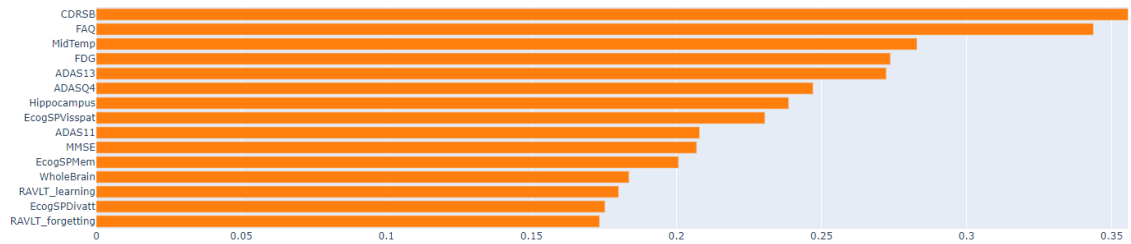
(a) Importancia global en el periodo de seguimiento de dos años



(b) Importancia global en el periodo de seguimiento de tres años

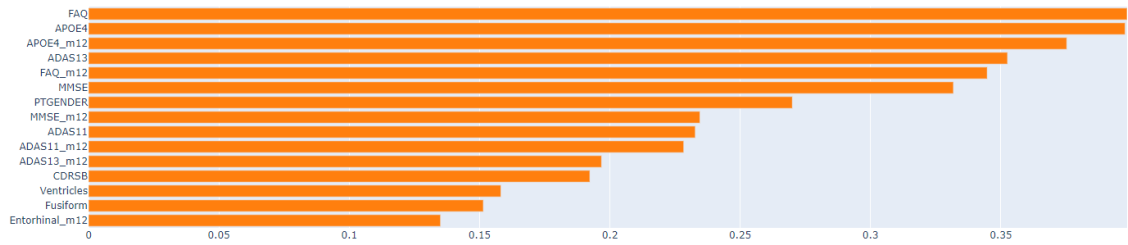


(c) Importancia global en el periodo de seguimiento de cuatro años

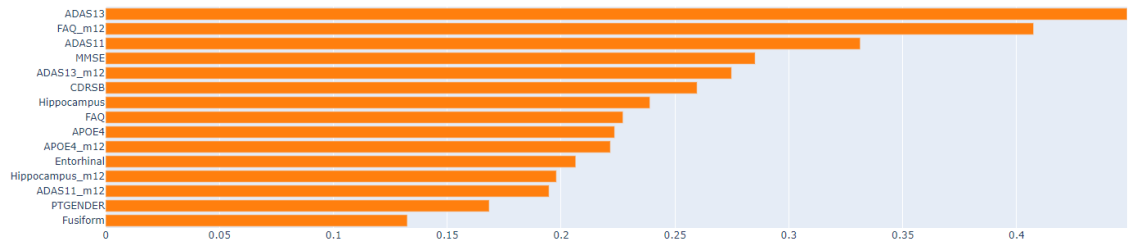


(d) Importancia global en el periodo de seguimiento de cinco años

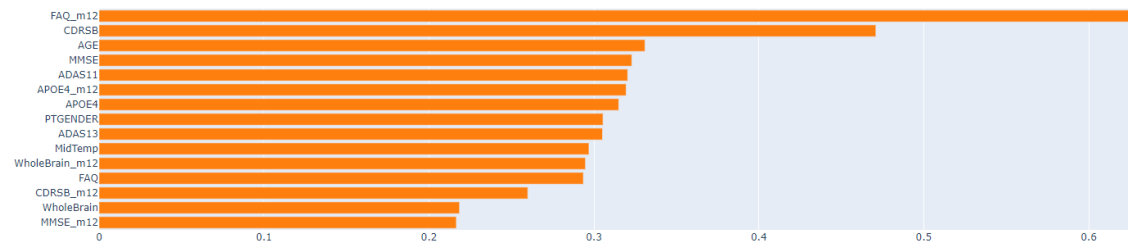
Figura 4: Modelo ADNIMERGE_set. Explicabilidad global en el mes 12



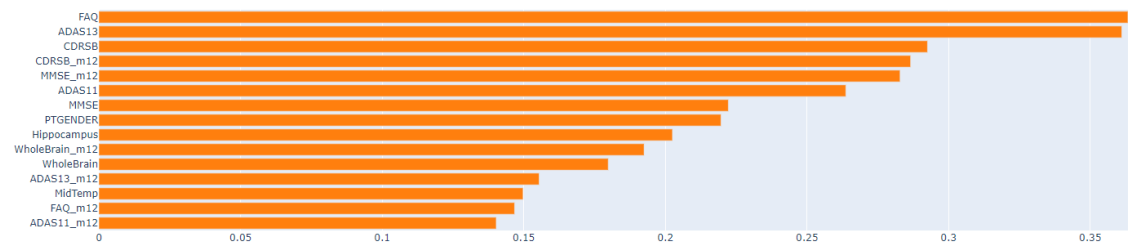
(a) Importancia global en el periodo de seguimiento de dos años



(b) Importancia global en el periodo de seguimiento de tres años



(c) Importancia global en el periodo de seguimiento de cuatro años



(d) Importancia global en el periodo de seguimiento de cinco años

Figura 5: Modelo EBM_set. Explicabilidad global utilizando dos instantes consecutivos

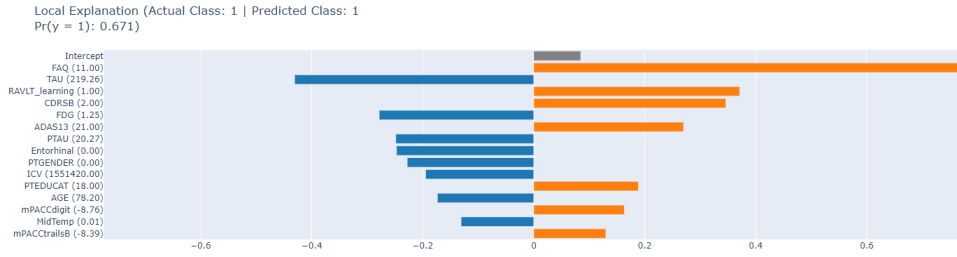
3.5. Explicabilidad local

Debido al importante número de pacientes, es inviable presentar los resultados de todos los pacientes uno por uno, por lo que se ha decidido seleccionar de manera aleatoria a cinco pacientes en el año 5. Se ha analizado la explicabilidad local para el mejor modelo de cada una de las pruebas realizadas, en el Anexo D se encuentran las figuras de la explicabilidad local para el resto de los modelos utilizados a lo largo del trabajo. Por otro lado, al igual que ocurre con la explicabilidad global, se ha simplificado el proceso analizando únicamente el quinto fold. De la misma manera, existe variabilidad entre los folds; sin embargo, esta no es muy significativa y se puede realizar un análisis fiable de cada uno de los modelos.

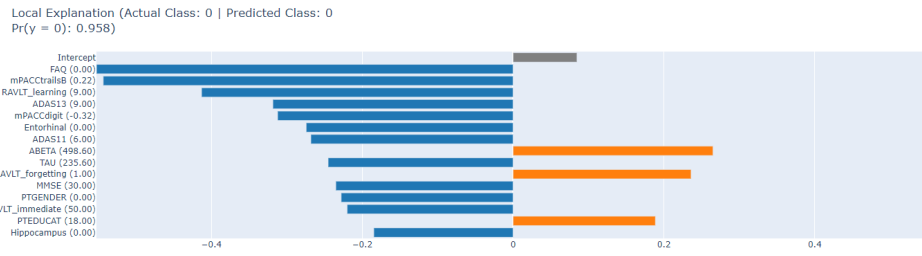
3.5.1. Explicabilidad local del modelo RSF_set para la visita en baseline

En el caso del modelo entrenado con RSF_set, de los cinco pacientes seleccionados, tres de ellos desarrollan Alzheimer mientras que los otros dos se mantienen estables con MCI. De estos cinco pacientes el sistema predice correctamente cuatro, como se puede observar en la Figura 6.

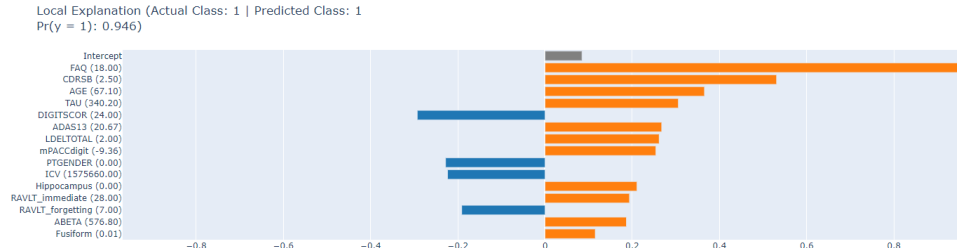
- Paciente 1: El paciente desarrolla Alzheimer, el sistema lo predice correctamente con una probabilidad del 67.1 %. Destacan con contribución positiva FAQ, RAVLT_learning y CDRSB, mientras que con contribución negativa TAU y FDG.
- Paciente 2: El paciente continua con MCI, el sistema lo predice correctamente con una probabilidad del 95.8 %. Las features más importantes, todas con contribución negativa son FAQ, mPACCtraillsB, RAVLT_learning, ADAS13 y mPACCdigit.
- Paciente 3: El paciente desarrolla Alzheimer, el sistema lo predice correctamente con una probabilidad del 94.6 %. Destacan todas las features más importantes con contribución positiva: FAQ, CDRSB, AGE, TAU y DIGITSCOR.
- Paciente 4: El paciente continua con MCI, el sistema lo predice correctamente con una probabilidad del 85.1 %. En este caso, las cuatro features más importantes presentan contribución negativa (TAU, FAQ, PTAU y PTEDUCAT) y con contribución positiva ADAS11.
- Paciente 5: El paciente desarrolla Alzheimer, el sistema falla ya que predice que continuará con MCI con una probabilidad del 92.8 %. Se tiene que las tres features más importantes (TAU, PTAU, FAQ) presentan contribución negativa y CDRSB con contribución positiva; asimismo, ADAS13 también tiene una contribución negativa.



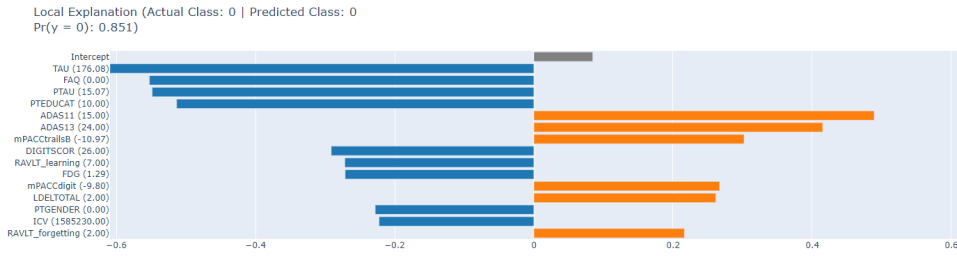
(a) Importancia de las features para el paciente 1



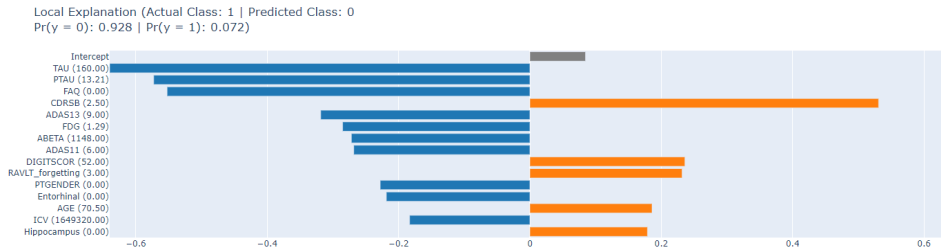
(b) Importancia de las features para el paciente 2



(c) Importancia de las features para el paciente 3



(d) Importancia de las features para el paciente 4



(e) Importancia de las features para el paciente 5

Figura 6: Modelo RSF_set. Explicabilidad local de 5 pacientes aleatorios

3.5.2. Explicabilidad local del modelo ADNIMERGE_set para la visita en el mes 12

Como se muestra en la Figura 7, el modelo acierta para los 5 pacientes seleccionados, de nuevo tres de ellos desarrollan Alzheimer mientras que los otros dos se mantienen estables con MCI.

- Paciente 1: El paciente continua con MCI, el sistema lo predice correctamente con una probabilidad del 96.2 %. Las features más importantes presentan contribución negativa, entre ellas destacan, CDRSB, MMSE, FAQ, EcogPtLang y EcogSPDivatt.
- Paciente 2: El paciente desarrolla Alzheimer, el sistema lo predice correctamente con una probabilidad del 82.1 %. En este caso se tiene que Hippocampus, MidTemp y mPACCdigit presentan contribución positiva y, CDRSB y EcogSPDivatt contribución negativa.
- Paciente 3: El paciente continua con MCI, el sistema lo predice correctamente con una probabilidad del 94.2 %. De nuevo las features más importantes presentan contribución negativa: CRDSB, FAQ, MidTemp, EcogSPVispat y EcogSPDivatt.
- Paciente 4: El paciente desarrolla Alzheimer, el sistema lo predice correctamente con una probabilidad del 82.0 %. Las tres features más importantes (Hippocampus, MidTemp y CDRSB) presentan contribución positiva. FAQ y EcogSPDivatt tienen contribución negativa.
- Paciente 5: El paciente desarrolla Alzheimer, el sistema lo predice correctamente con una probabilidad del 79.5 %. ADAS13 MidTep WholeBrain y FDG tienen contribución positiva mientras que CDRSB tiene contribución negativa.

3.5.3. Explicabilidad local del modelo EBM_set utilizando dos instantes temporales consecutivos

De la misma forma que en el apartado anterior, el modelo acierta para los 5 pacientes seleccionados. La Figura 8 muestra que, en este caso, dos de ellos desarrollan Alzheimer mientras que los otros tres se mantienen estables con MCI.

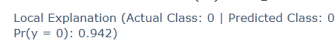
- Paciente 1: El paciente continua con MCI, el sistema lo predice correctamente con una probabilidad del 97.3 %. Las features más importantes tienen contribución negativa: MMSE en el mes 12, CDRSB, ADAS13, ADAS11 y FAQ.
- Paciente 2: El paciente continua con MCI, el sistema lo predice correctamente con una probabilidad del 99.0 %. Igual que en el paciente 1, las features más importantes tienen contribución positiva: MMSE en el mes 12, CDRSB, FAQ, ADAS13 y MMSE en baseline .
- Paciente 3: El paciente continua con MCI, el sistema lo predice correctamente con una probabilidad del 56.6 %. En este caso, la feature más relevante, FAQ, tiene contribución negativa mientras que MMSE en el mes 12, ADAS13 en baseline y el mes 12 y ADAS11 tienen contribución positiva.
- Paciente 4: El paciente desarrolla Alzheimer, el sistema lo predice correctamente con una probabilidad del 91.7 %. ADAS13, MidTemp en baseline y el mes 12, Hippocampus y ADAS11 tienen contribución positiva.
- Paciente 5: El paciente desarrolla Alzheimer, el sistema lo predice correctamente con una probabilidad del 84.7 %. La segunda feature más relevante, CDRSB, tiene contribución negativa, mientras que WholeBrain, ADAS13, ADAS11 y MidTemp tienen contribución positiva.



(a) Importancia de las features para el paciente 1



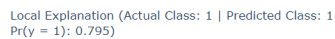
(b) Importancia de las features para el paciente 2



(c) Importancia de las features para el paciente 3

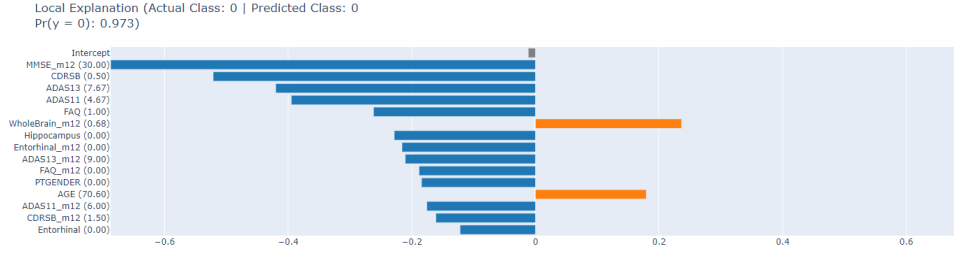


(d) Importancia de las features para el paciente 4

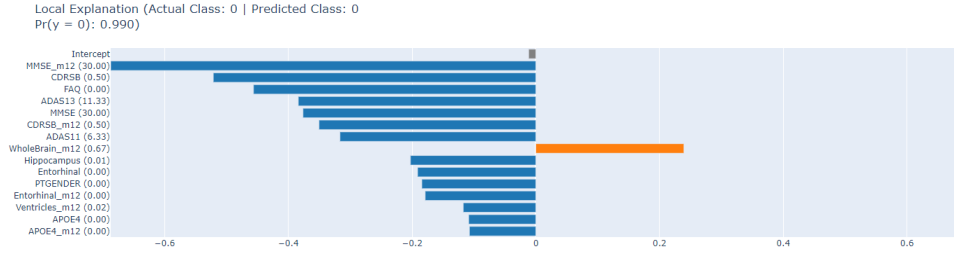


(e) Importancia de las features para el paciente 5

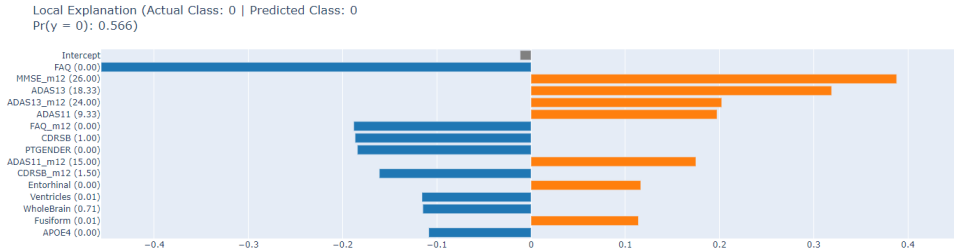
Figura 7: Modelo ADNIMERGE_{set}. Explicabilidad local de 5 pacientes aleatorios



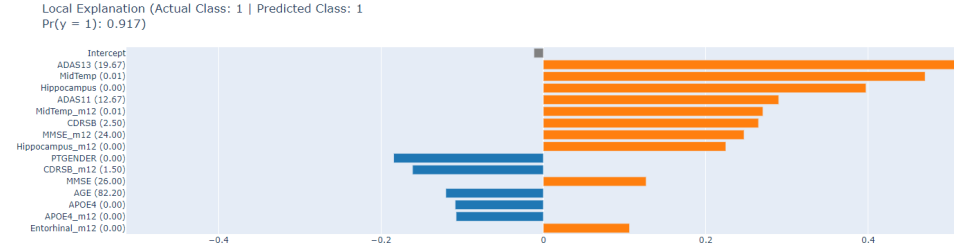
(a) Importancia de las features para el paciente 1



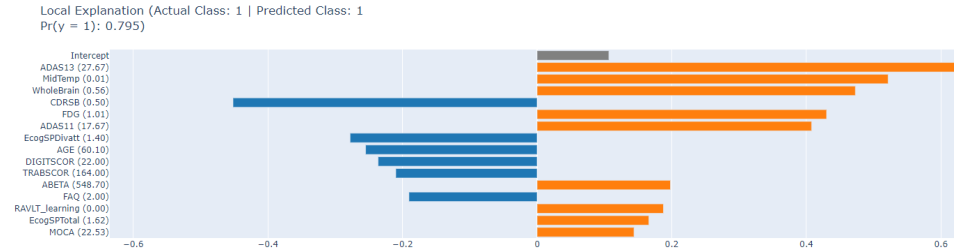
(b) Importancia de las features para el paciente 2



(c) Importancia de las features para el paciente 3



(d) Importancia de las features para el paciente 4



(e) Importancia de las features para el paciente 5

Figura 8: Modelo EBM_set. Explicabilidad local de 5 pacientes aleatorios

4. Discusión

La detección temprana de la enfermedad del Alzheimer es de crucial importancia para detener su progreso en las etapas iniciales y mejorar la calidad de vida de los pacientes y sus familiares. El trabajo realizado en este TFG se basa en el estudio de modelos explicables de EBM que utilizan features de distintos tipos para predecir si los pacientes con MCI desarrollarán AD durante un período de seguimiento con una longitud de cinco años.

En este estudio, se han comparado modelos que utilizan combinaciones de diferentes longitudes temporales y diferentes conjuntos de features. Se encuentra que, el modelo entrenado con RSF_set muestra un rendimiento superior cuando se utilizan los datos de la primera visita del estudio para realizar la predicción. Por otro lado, al utilizar los datos de los pacientes en su segunda visita (en el mes 12), el modelo con mejor rendimiento es el entrenado con ADNIMERGE_set. Finalmente, cuando se incluyen dos instantes temporales consecutivos en los conjuntos de features, se concluye que el modelo con mejor rendimiento es el entrenado con EBM_set. Estos resultados demuestran que el rendimiento del modelo es altamente dependiente del conjunto de features utilizado en cada prueba.

El modelo con mejor rendimiento de entre todas las pruebas realizadas es el modelo entrenado con ADNIMERGE_set cuando se utilizan los datos de la segunda visita para realizar la predicción. Dicho modelo presenta prestaciones bastante positivas en términos generales a pesar de su menor rendimiento para las métricas sensitivity y f1-score. No obstante, este modelo ha alcanzado un 89.11 % en promedio en AUC. Este valor de AUC refleja la capacidad del modelo para distinguir correctamente entre las clases positivas y negativas. Este modelo obtiene un resultado superior en AUC en comparación al mejor modelo utilizando los datos de la visita en baseline y al mejor modelo utilizando dos instantes temporales consecutivos.

Por otro lado, en el caso del modelo entrenado con EBM_set que utiliza los datos de la primera visita del estudio, se observa un descenso generalizado en el rendimiento en todas las métricas en el año 5. Este resultado difiere con los resultados obtenidos en el artículo en el que se basa este trabajo [14], donde el rendimiento más bajo se obtenía en los años 2 y 3, seguido de una mejora en los dos últimos años alcanzando el mejor resultado en el quinto año, como se muestra en la Figura 9. Este aspecto de discordancia entre los resultados es interesante y se puede deber, principalmente, al hecho de que en este trabajo se ha decidido eliminar la feature DX_bl del conjunto de features (EBM_set) utilizado para entrenar al modelo. Como se ha mencionado anteriormente, esta feature presenta una alta correlación con la variable objetivo y puede distorsionar la predicción del modelo. Asimismo, también existen divergencias entre los resultados de explicabilidad global. Se destaca que no se observa tanta variabilidad en la importancia de las features entre los distintos años, en otras palabras, la importancia de las features no cambia significativamente dependiendo del año, a diferencia del artículo original. Además, en este caso, la importancia de las features anatómicas es bastante inferior en comparación con el estudio realizado en el artículo, donde son las más influyentes en la mayor parte de los años.

Asimismo, la eliminación de la feature DX_bl de los conjuntos que la incluían supuso una mejora notable en los resultados en sensitivity y f1-score para el primer año en todos los modelos que utilizan los datos en baseline para realizar la predicción, por lo que todos los experimentos posteriores se hicieron obviando esta feature. La presencia de esta feature disminuía la capacidad del modelo para identificar los casos positivos correctamente en el primer año del estudio.

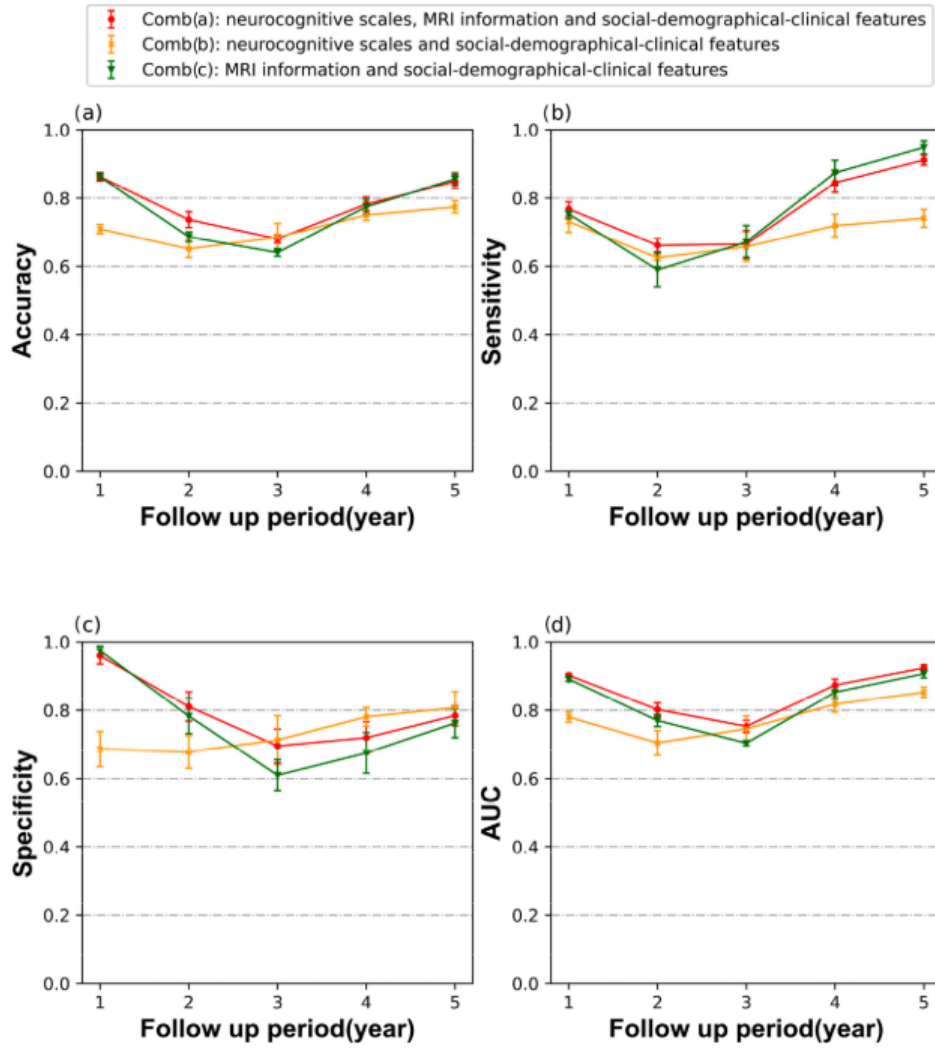


Figura 9: Resultados obtenidos en el artículo de referencia

Por otra parte, el bajo rendimiento del modelo entrenado con ODE-RGRU_set se puede deber a varios factores. En primer lugar, este conjunto incluye el menor número de features, con solo 9 en total, de las cuales 6 corresponden a medidas MRI y las otras 3 a medidas de evaluación neuropsicológica. Pese a que estas son features aparentemente influyentes para la predicción, el análisis de la explicabilidad realizado muestra la escasa relevancia de las features anatómicas en comparación con otro tipo de features. Por tanto, esto unido a la falta de otras features bastante relevantes como APOE ϵ 4, un biomarcador que representa el factor genético de riesgo más potente conocido para los casos de AD tardío [17], podrían provocar la disminución de rendimiento del modelo. Asimismo, la ausencia de FAQ, la cual presenta especial relevancia en los modelos como se ha observado en el análisis de explicabilidad global, también podría influir negativamente en el rendimiento. FAQ es una escala de valoración breve basada en informantes para la medición indirecta de las actividades instrumentales de la vida diaria en adultos mayores. Muestra una alta validez discriminativa para la diferenciación de la independencia funcional en pacientes con demencia y MCI, y una alta validez predictiva en la detección de aquellos pacientes con riesgo de progresión de MCI a AD [18].

El análisis de los resultados utilizando los datos de la visita en el mes 12 para realizar la predicción, revela una ligera mejora en los resultados para todos los modelos, especialmente notable en los modelos entrenados con ODE-RGRU_set y ADNIMERGE_set en comparación con los resultados utilizando los datos de baseline. La mejora significativa observada en el modelo entrenado con ODE-RGRU_set al utilizar los datos de la segunda visita para la predicción, en contraste con los datos de la visita inicial, podría atribuirse a la naturaleza progresiva del Alzheimer. Los efectos de esta enfermedad pueden no ser claramente evidentes durante el primer año, y por tanto no detectarse en las medidas de evaluación neuropsicológica, medidas que son las más relevantes para la predicción en este conjunto. Por lo que el uso de datos de la segunda visita permite una manifestación más clara de la enfermedad en dichas medidas, mejorando de esta forma el rendimiento del modelo. Concluyendo con la visita en el mes 12, se puede destacar una mejora significativa en los resultados de sensitivity en todos los modelos en comparación con los resultados de la primera visita. Por tanto, se puede concluir que la reducción en el periodo de seguimiento y por tanto la proximidad a la variable objetivo, mejora los resultados para esta métrica.

Por otro lado, la utilización de conjuntos que incluyen dos instantes temporales consecutivos no supone una mejora generalizada en los resultados, sino que es muy dependiente del modelo. El modelo entrenado con ODE-RGRU_set se favorece de la inclusión de dos instantes temporales consecutivos mejorando las prestaciones en comparación con la visita en el mes 12. De manera similar, el modelo entrenado con EBM_set también mejora los resultados en comparación a la visita en baseline y el mes 12. Lo más destacable es la reducción en el rendimiento del modelo entrenado con ADNIMERGE_set, ya que presenta peores resultados en comparación con la visita inicial y el mes 12, donde era el mejor modelo.

El estudio realizado en este trabajo aporta pruebas, desde la perspectiva del aprendizaje automático, del importante papel de FAQ y el biomarcador APOE ϵ 4 en la predicción de la enfermedad del Alzheimer, que, como se ha comentado a lo largo de esta memoria, son factores muy relevantes en el diagnóstico de la enfermedad. Asimismo, en el caso de las features de medidas MRI, se ha destacado que su papel en la predicción no es tan importante como cabría esperar, siendo las más relevantes Entorhinal y posteriormente MidTemp.

Por otra parte, en este estudio se ha mostrado la gran importancia de la feature FDG en la predicción, aunque esta importancia disminuye cuando se utilizan los datos de la visita en el mes 12 y cuando se utilizan dos instantes temporales consecutivos. Asimismo, se ha demostrado en varios estudios que se trata de una herramienta adecuada para apoyar el diagnóstico temprano de la demencia ya que es clave en el diagnóstico de la enfermedad del Alzheimer, específicamente señalando el hipometabolismo en la corteza cingulada posterior y parietotemporal, y en los lóbulos frontales y temporales mediales, además de permitir la caracterización del MCI [19]. De igual manera, la proteína TAU, presenta un amplio papel patológico en varias enfermedades neurodegenerativas, y por tanto en el Alzheimer, ya que la hiperfosforilación anormal de TAU desempeña funciones patológicas perjudiciales que, en última instancia, conducen a la neurodegeneración [20].

Sin embargo, en relación a los conjuntos MinimalRNN_set, RSF_set y ADNIMERGE_set, que incluyen la proteína TAU como feature, sorprende la escasa influencia de esta feature en la predicción del modelo en términos generales. A pesar de su rol importante en las enfermedades neurodegenerativas, en los tres conjuntos esta proteína no tiene una gran relevancia. Destaca solo en los años 2 y 5 para RSF_set en baseline y no es nada relevante en ADNIMERGE_set en el caso de la visita en el mes 12. Esta escasa relevancia puede deberse a que la proporción de valores faltantes ronda el 50 %, lo que puede afectar negativamente al modelo e impide que se beneficie de la presencia de esta feature.

Continuando con las features con un gran número de valores faltantes, es preciso señalar que gracias al análisis de la explicabilidad global, se ha podido comprobar que aquellas features con un gran número de valores faltantes apenas son relevantes para la predicción en términos generales. Por tanto, el riesgo de que este factor afecte negativamente al modelo, debido a que la imputación se realiza en base a aproximadamente la mitad de los datos disponibles, es muy bajo.

Por otro lado, el CDRSB presenta una gran relevancia en la predicción, también se ha demostrado en varios estudios como una herramienta esencial en la evaluación y seguimiento de la progresión de la enfermedad de MCI a AD. Esta feature no solo permite una evaluación precisa de la severidad inicial de la enfermedad, sino que también muestra que las tasas de progresión varían significativamente con la severidad de la misma, lo que es fundamental para este modelo en concreto el cuál trata de predecir la progresión de un paciente con MCI a AD [21].

Finalmente, se puede destacar la importancia de RAVLT_immediate en la explicabilidad de los tres primeros años del estudio (o los años 2 y 3 si se utilizan los datos de la visita en el mes 12). Esta feature en concreto presenta asociación con la atrofia cerebral relacionada con el AD, y tiene una mayor precisión predictiva para la predicción de la conversión a AD en pacientes con MCI [22]. No obstante, en los dos últimos años su relevancia disminuye bastante. Esta disminución puede deberse a que la atrofia cerebral se da en mayor medida en pacientes con AD temprano en comparación a los pacientes con AD tardío [23], lo que disminuye la relevancia de esta feature en las predicciones de los años finales. Asimismo, en estos dos años del estudio la muestra de pacientes disminuye considerablemente; además hay una proporción similar (60-40) de pacientes con MCI y con AD. Por tanto, esta mayor presencia de pacientes con AD podría también afectar a la relevancia de esta feature en la predicción.

Al analizar la explicabilidad global en los modelos que utilizan los datos de la visita en el mes 12 para realizar la predicción, se ha observado una menor influencia de las features anatómicas a lo largo del estudio. Estas solo destacan en los dos últimos años, en contraste con los modelos que utilizaban los datos de baseline, donde algunas features anatómicas, como el Entorhinal, presentaban una relevancia media en prácticamente todos los años. Además, se ha observado una mayor variación en la importancia de las features dependiendo del año. El año 4 es particularmente notable por los mayores cambios en importancia, con features que apenas tenían relevancia en los dos años anteriores cobrando especial importancia. Sin embargo, las features que presentaban la mayor importancia en la visita inicial también son las más relevantes en la visita en el mes 12.

Este análisis también resalta la importancia de mPACCdigit en los dos primeros años (años 2 y 3). El Preclinical Alzheimer Cognitive Composite (PACC) es una métrica que se calcula a partir de las puntuaciones de varios tests, y en el caso de mPACCdigit, incluye el test específico de Digit Symbol Substitution [24]. El mPACC está relacionado con la memoria de trabajo. Las etapas tempranas del Alzheimer se caracterizan por la disfunción ejecutiva y el deterioro de la memoria de trabajo, además de déficits en la memoria episódica. Estos déficits cognitivos comienzan cuando el paciente presenta MCI y son un indicio de progresión hacia el AD. Por tanto, el seguimiento de la memoria de trabajo a lo largo del tiempo puede señalar etapas de declive cognitivo [25]. El inicio de estos déficits cognitivos en las etapas tempranas del Alzheimer, junto con el uso de datos de los pacientes en la visita del mes 12 para la predicción, explican la importancia del mPACCdigit en los dos primeros años del estudio realizado en el apartado 3.4.2. Además, esto también explica por qué la relevancia de esta feature no fue destacable cuando se utilizaban los datos de la primera visita para la predicción.

Por otro lado, del análisis de la explicabilidad global cuando se utilizan dos instantes temporales consecutivos, se puede destacar que las features de la segunda visita son principalmente relevantes en el último año y en el primero (año 2), aunque en este un poco menos relevantes. Esta escasa relevancia podría indicar por qué la mejora en los resultados no es general cuando se incluyen dos instantes temporales, sino que depende mucho del conjunto de features utilizado. Asimismo, al igual que ocurría en la explicabilidad del mes 12 las features anatómicas presentan escasa relevancia, igualmente existe bastante variación en la importancia de las features entre los años (aunque esta es menor), pero las features más relevantes continúan siendo prácticamente las mismas que en el caso de baseline y el mes 12 en términos generales.

En cuanto a la explicabilidad local, a pesar de la dificultad de obtener conclusiones generales debido a la heterogeneidad de las características de los pacientes, es preciso señalar que las features anatómicas tienden a tener una contribución positiva en la predicción, aunque de igual modo que ocurría en la explicabilidad global su relevancia tiende a ser baja. Además, se observa una diferencia en las features consideradas cuando el sistema falla en las predicciones. Si el sistema predice incorrectamente el desarrollo de Alzheimer en un paciente que en realidad sigue siendo MCI, las features más relevantes son las anatómicas como Ventricles, WholeBrain, o Entorhinal, además de otras como MMSE y ADAS11. Por otro lado, si el sistema predice que el paciente seguirá estable pero en realidad desarrolla AD, las características más recurrentes con una contribución negativa suelen ser MMSE, CDRSB, PTGENDER, APOE ϵ 4 y ADAS11.

Para finalizar, este trabajo presenta el punto fuerte de que se centra en la explicabilidad ya que, aunque se han realizado varios estudios de investigación para examinar la enfermedad del Alzheimer, y concretamente la conversión de MCI a AD [26], su enfoque principal se centra en la precisión de los algoritmos de aprendizaje automático de referencia y no tanto en la explicabilidad. En este estudio no sólo se busca lograr buenos resultados, sino que también se da valor a la explicabilidad, incorporando técnicas de explicabilidad tanto globales como locales. De esta forma se entiende que a pesar de que el modelo con mejores prestaciones presente un menor rendimiento para las métricas sensitivity y f1-score, la particularidad de que sea un modelo explicable unido a su valor alto en AUC, lo hacen candidato para su aplicación en el ámbito médico y que tenga un impacto positivo en la predicción de MCI a AD.

5. Conclusiones

En este trabajo se ha utilizado el método EBM para predecir la conversión de pacientes con MCI a AD en diferentes periodos de seguimiento utilizando la información de la base de datos ADNI. Asimismo, no sólo se han presentado los resultados obtenidos con dicho método sino que también se ha expuesto la explicabilidad del mismo, tanto global como local. Se han realizado diversas pruebas con distintos conjuntos de features para conseguir el mejor rendimiento utilizando este método. Se ha concluido que el modelo con mejor rendimiento fue el modelo entrenado con `ADNIMERGE_set` y que utiliza los datos de la segunda visita para realizar la predicción, de manera que se tenía un período de seguimiento desde el año 2 hasta el año 5.

El método utilizado se encuentra accesible fácilmente ya que se trata de un método de una librería de Python, lo que unido a su naturaleza glass-box, lo convierten en un candidato ideal para la aplicación práctica en el ámbito de la medicina. La aplicación de este método en el campo de la enfermedad del Alzheimer en concreto puede ser útil para el diagnóstico precoz con el fin de frenar el deterioro cognitivo y retrasar la aparición de la demencia. Además se ha puesto el foco sobre aquellas features o marcadores más influyentes para la predicción de la conversión de MCI a AD cuando se utiliza este método en concreto, y se han contrastado con otros estudios realizados en el campo de la enfermedad del Alzheimer, validando su importancia en el diagnóstico de esta enfermedad.

En el futuro, se tiene previsto ahondar en mayor medida en la selección de features, partiendo del modelo con mejores resultados, de manera que se pueda localizar o determinar aquellas features que pueden estar impidiendo un rendimiento superior del modelo o incluir otras que puedan mejorar el rendimiento. Continuando con este aspecto, también se quiere analizar el comportamiento del mejor modelo utilizando interacciones entre las features, ya que en este trabajo, desde el principio se decidió no utilizar interacciones, las cuales pueden mejorar la precisión del modelo. Por otro lado, también se pretende investigar el motivo por el cual el modelo presenta un rendimiento inferior en sensitivity y f1-score.

Finalmente, dado que el Alzheimer es un tipo de demencia, se podría aplicar este modelo para la investigación de otros tipos de neurodegeneración, como por ejemplo el Parkinson, con el objetivo de identificar posibles features relacionadas entre sí para ampliar el conocimiento acerca de este tipo de enfermedades. En conclusión, este trabajo no solo ha demostrado la viabilidad del método EBM para la predicción de la conversión de MCI a AD, sino que también ha resaltado la importancia de la explicabilidad en sistemas de inteligencia artificial en el ámbito médico. Esto no solo facilita la confianza y aceptación por parte de los profesionales de la salud, sino que también asegura que las decisiones del modelo sean transparentes y basadas en datos sólidos.

Referencias

- [1] Anil Kumar, Jaskirat Sidhu, Amandeep Goyal, and Jack W. Tsao. *Alzheimer Disease*. StatPearls Publishing, Treasure Island (FL), 2023. [View Article](#).
- [2] Philip Scheltens, Bart De Strooper, Miia Kivipelto, Henne Holstege, Gael Chételat, Charlotte E Teunissen, Jeffrey Cummings, and Wiesje M van der Flier. “Alzheimer’s disease”. In: *The Lancet* 397(10284):1577-1590 (2021). [View Article](#).
- [3] P. A. Boyle, R. S. Wilson, N. T. Aggarwal, Y. Tang, and D. A. Bennett. “Mild cognitive impairment, Risk of Alzheimer disease and rate of cognitive decline”. In: *Neurology* 67(3):441-44 (2006). [View Article](#).
- [4] Roch A. Nianogo, Amy Rosenwohl-Mack, Kristine Yaffe, Anna Carrasco, Coles M. Hoffmann, and Deborah E. Barnes. “Risk Factors Associated With Alzheimer Disease and Related Dementias by Sex and Race and Ethnicity in the US”. In: *JAMA Neurology* 79(6):584-591 (2022). [View Article](#).
- [5] X.-X. Zhang, Y. Tian, Z.-T. Wang, Y.-H. Ma, Lan Tan, and Jin-Tai Yu. “The Epidemiology of Alzheimer’s Disease Modifiable Risk Factors and Prevention”. In: *J Prev Alzheimers Dis* 8:313-321 (2021). [View Article](#).
- [6] Afzal Sitaram, Maqsood Muazzam, Khan, Umair, Mehmood Irfan, Nawaz Hina, Aadil Farhan, Song Oh-Young, and Yunyoung Nam. “Alzheimer Disease Detection Techniques and Methods: A Review”. In: *International Journal of Interactive Multimedia and Artificial Intelligence* 6(7):26-38 (2021). [View Article](#).
- [7] Sajid Ali, Tamer Abuhmed, Shaker El-Sappagh, Khan Muhammad, Jose M. Alonso-Moral, Roberto Confalonieri, Riccardo Guidotti, Javier Del Ser, Natalia Díaz-Rodríguez, and Francisco Herrera. “Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence”. In: *Information Fusion* 99:101805 (2023). [View Article](#).
- [8] Rai and Arun. “Explainable AI: from black box to glass box”. In: *J. of the Acad. Mark. Sci.* 48:137-141 (Jan 2020). [View Article](#).
- [9] Leo Breiman. “Bagging predictors”. In: *Mach Learn* 24:123-140 (1996). [View Article](#).
- [10] Jerome H. Friedman. “Greedy function approximation: A gradient boosting machine.” In: *The Annals of Statistics* 29(5):1189-1232 (2001). [View Article](#).
- [11] Jerome H. Friedman. “Stochastic gradient boosting”. In: *Computational Statistics Data Analysis* 38(4):367-378 (2002). [View Article](#).
- [12] S. Jeong, W. Jung, J. Sohn, and H. I. Suk. “Deep geometrical learning for Alzheimer’s disease progression modeling”. In: *2022 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2022, pp. 211-220. [View Article](#).
- [13] Minh Nguyen, Tong He, LiAn, Daniel C Alexander, Jiashi Feng, and Alzheimer’s Disease Neuroimaging Initiative B T Thomas Yeo and. “Predicting Alzheimer’s disease progression using deep recurrent neural networks”. In: *NeuroImage* 222:117203 (2022). [View Article](#).
- [14] Jiaxin Cai, Weiwei Hu, Jiaojiao Ma, Aima Si, Shiyu Chen, Lingmin Gong, Yong Zhang, Hong Yan, Fangyao Chen, and Alzheimer’s Disease Neuroimaging Initiative. “Explainable Machine Learning with Pairwise Interactions for Predicting Conversion from Mild Cognitive Impairment to Alzheimer’s Disease Utilizing Multi-Modalities Data”. In: *Brain Sci.* 13(11):1535 (2023). [View Article](#).

- [15] Alessia Sarica, Federica Aracri, Maria Giovanna Bianco, Fulvia Arcuri, Andrea Quattrone, Aldo Quattrone, and for the Alzheimer’s Disease Neuroimaging Initiative. “Explainability of random survival forests in predicting conversion risk from mild cognitive impairment to Alzheimer’s disease”. In: *Brain Inf.* 10:31 (2023). [View Article](#).
- [16] Olga Voevodskaya, Andrew Simmons, Richard Nordenskjöld, Joel Kullberg, Håkan Ahlström, Lars Lind, Lars-Olof Wahlund, Elna-Marie Larsson, Eric Westman, and Alzheimer’s Disease Neuroimaging Initiative. “The effects of intracranial volume adjustment approaches on multiple regional MRI volumes in healthy aging and Alzheimer’s disease”. In: *Front. Aging Neurosci.* 6:264 (2014). [View Article](#).
- [17] Holly C. Hunsberger, Priyanka D. Pinky, Warren Smith, Vishnu Suppiramaniam, and Miranda N. Reed. “The role of APOE4 in Alzheimer’s disease: strategies for future therapeutic interventions”. In: *Neuronal Signaling* 3(2):NS20180203 (2019). [View Article](#).
- [18] Ondrej Bezdicek. “Chapter 27 - The functional activities questionnaire: applications to aging”. In: *Assessments, Treatments and Modeling in Aging and Neurological Disease*. Ed. by Colin R. Martin, Victor R. Preedy, and Rajkumar Rajendram. Academic Press, 2021, pp. 293–303. ISBN: 978-0-12-818000-6. [View Article](#).
- [19] Lisa Mosconi et al. “Multicenter Standardized 18F-FDG PET Diagnosis of Mild Cognitive Impairment, Alzheimer’s Disease, and Other Dementias”. In: *Journal of Nuclear Medicine* 49(3):390-398 (2008). [View Article](#).
- [20] Shibi Muralidar, Senthil Visaga Ambi, Saravanan Sekaran, Diraviyam Thirumalai, Balamurugan Palaniappan Shibi Muralidar, Senthil Visaga Ambi, Saravanan Sekaran, Diraviyam Thirumalai, and Balamurugan Palaniappan. “Role of tau protein in Alzheimer’s disease: The prime pathological player”. In: *International Journal of Biological Macromolecules* 163:1599-1617 (2020). [View Article](#).
- [21] Mahesh N Samtani, Nandini Raghavan, Gerald Novak, Partha Nandy, Vaibhav A Narayan, and Alzheimer’s Disease Neuroimaging Initiative. “Disease progression model for Clinical Dementia Rating–Sum of Boxes in mild cognitive impairment and Alzheimer’s subjects from the Alzheimer’s Disease Neuroimaging Initiative”. In: *Neuropsychiatric Disease and Treatment* 10:929-952 (2014). [View Article](#).
- [22] Elahesh Moradi, Ilona Hallikainen, Tuomo Hänninen, Jussi Tohka, and Alzheimer’s Disease Neuroimaging Initiative. “Rey’s Auditory Verbal Learning Test scores can be predicted from whole brain MRI in Alzheimer’s disease”. In: *NeuroImage: Clinical* 13:415-427 (2017). [View Article](#).
- [23] C. R. Jack et al. “Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD”. In: *Neurology* 62(4):591-600 (2004). [View Article](#).
- [24] Michael C. Donohue, Reisa A. Sperling, Ronald Petersen and Chung-Kai Sun, MS Michael W. Weiner, Paul S. Aisen, and for the Alzheimer’s Disease Neuroimaging Initiative. “Association Between Elevated Brain Amyloid and Subsequent Cognitive Decline Among Cognitively Normal Persons”. In: *JAMA* 317(22):2305–2316 (2017). [View Article](#).
- [25] Anna-Mariya Kirova, Rebecca B. Bays, and Sarita Lagalwar. “Working Memory and Executive Function Decline across Normal Aging, Mild Cognitive Impairment, and Alzheimer’s Disease”. In: *BioMed Research International* 2015(1):748212 (2015). [View Article](#).
- [26] El-Sappagh S, Saleh H, Ali F, Amer E, and Abuhmed T. “Two-stage deep learning model for Alzheimer’s disease detection and prediction of the mild cognitive impairment time”. In: *Neural Comput Applic* 34:14487–14509 (2022). [View Article](#).